

Manas Tokale

manastokale@gmail.com | 447 902 6908 | Champaign, IL (Relocation Ready) | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Machine Learning Engineer and Data Scientist with 4+ years of experience building applied AI, forecasting, retrieval, and computer vision systems for research and enterprise workflows. Strong record of shipping decision-facing products, improving operational speed and quality, and translating complex models into deployable systems with measurable business impact.

EDUCATION

University of Illinois, Urbana-Champaign | Champaign, Illinois **August 2024 - May 2025**

Master of Science in Business Analytics (Data Science Track) (STEM)

- *Advanced Data Science, Big Data Infrastructures, Enterprise Database Management, Marketing Analytics, Business Intelligence*

NMIMS University | Mumbai, India **July 2017 - August 2021**

Bachelor of Technology in Data Science

- *Artificial Intelligence, Machine Learning, Natural Language Processing, Computer Vision, Cloud Computing*

TECHNICAL SKILLS

Languages : Python, SQL, R, JavaScript, C++.

Applied AI / ML : scikit-learn, TensorFlow, PyTorch, XGBoost, OpenCV, Transformers, Causal Inference, A/B Testing, SHAP.

Generative AI : LLM, SLM, RAG, MCP, Fine-Tuning, QLoRA, LangChain, LangGraph, Google ADK.

Cloud / MLOps : AWS ([CLF-C02](#)), GCP ([BigQuery](#)), Azure, Docker, Git, CI/CD, MLFlow.

Databases & Tools : MongoDB, DuckDB, PySpark, Tableau, Power BI, Firebase, Shell Scripting, CUDA, Postman.

WORK EXPERIENCE

Data Science Researcher | **University of Illinois, Urbana-Champaign** | Champaign, Illinois **October 2024 - Ongoing**

Tech Stack : Python, Mathematics, Hypothesis Testing, Git, Cloud Computing, Kanban, Google ADK

- Engineered reusable Python and LLM workflows for resume screening, workload tracking, and data capture, accelerating review throughput 10x while reducing research timelines 30% and operating cost 40% for admissions and research operations.
- For entrepreneurship and GenAI evaluation workflows, built agentic analytics tooling that improved project triage, standardized evidence capture, and made faculty-facing decisions faster and more auditable.

Data Scientist | **CNeutral** | Champaign, Illinois **September 2024 - May 2025**

Tech Stack : Python, PyMuPDF, LLM, ChromaDB, FAISS

- Scaled an NLP extraction pipeline for a 9,000-firm sovereign fund, improving data quality by 90% via prompt engineering.
- Designed a RAG-style ESG retrieval system with evidence-linked results, cutting analysis time by 40% and making downstream review more defensible.

Machine Learning Engineer | **Technokart Consultancy Services** | Mumbai, India **June 2020 - July 2023**

Tech Stack : Python, AWS, Linux (RHEL), PaddleOCR, SDLC, TIFF

- Deployed fine-tuned YOLO models for **PII redaction** in an eKYC pipeline, achieving 92% recall and ensuring compliance.
- Slashed audit overhead by 60% via a multilingual OpenAI Whisper & DistilBERT pipeline for **automated script compliance**.
- Optimized inference latency by 40% (10s to 6s) for a microservice processing 1M+ documents daily, maintaining 96% uptime.
- Eliminated critical deployment failures by establishing CI/CD pipelines and peer reviews in an Agile environment.

PROJECTS

Lineage - Screenplay Intelligence | [Github](#)

Tech Stack : Python, RAG, RRF, ChromaDB, Redis, Gemini API, Tailwind, Vite

- Solved continuity drift in narrative chat, by building a deployable screenplay-aware RAG workspace with time-bounded retrieval, reranking, and relationship memory, turning flat transcripts into a grounded AI product.

LocalPsych - Privacy-Preserving Mental Health AI | [Github](#)

Tech Stack : QLoRA, Gemma 4 E4B, HuggingFace PEFT/TRL, bitsandbytes

- Fine-tuned Gemma 4 via QLoRA** on a curated dataset of 8,861 empathetic interactions, engineering a fully on-device, privacy-preserving mental health assistant with integrated harmful-content filtering.

Fraud Detection, Behavioral Risk Modeling and Macroeconomic Analysis on Customer Banking Data | [Github](#)

Tech Stack : Python, Time-Series Forecasting, Fraud & Risk Analytics

- Developed a time-aware predictive model to forecast spending (**95% accuracy**) and detect fraud (**87% recall**).
- Utilized regression analysis to determine the statistical significance between consumer spending and macroeconomic factors.

Large Scale Forecasting for System Maintenance | [INFORMS RAS \(Ranked 7th\)](#) | [Github](#)

Tech Stack : Python, DuckDB SQL, Scikit-learn, Pandas, XGBoost, Feature Engineering, Time-Series Data

- Built an **XGBoost**, **CatBoost**, and **LightGBM** GPU-accelerated ensemble, ranking 6th globally by improving failure prediction accuracy by **19%** and accelerating **10GB+** sensor data processing speeds by 5x through optimized pipelines.

Contact Center Forecasting | [Github](#)

Tech Stack : XGBoost, Circular Kernel Smoothing

- Engineered a hybrid top down forecasting architecture by combining a lag-featured XGBoost regressor with a kernel-smoothed statistical distribution model that finished Top-7 with a 15.42 composite score by solving slot-level data quality and demand-shaping issues.